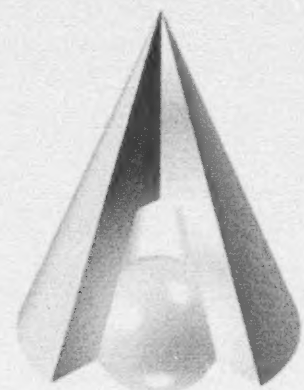
Catalogue no. 11F0019M — No. 362
ISSN 1205-9153
ISBN 978-1-100-25032-8

Analytical Studies Branch Research Paper Series

# The Effects of Cancer on Employment and Earnings of Cancer Survivors



by Sung-Hee Jeon

Release date: September 30, 2014

Statistics Canada Statistique Canada

Canada

## How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

**email** at infostats@statcan.gc.ca,

**telephone**, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

| | |
|---|---|
| • Statistical Information Service | 1-800-263-1136 |
| • National telecommunications device for the hearing impaired | 1-800-363-7629 |
| • Fax line | 1-877-287-4369 |

**Depository Services Program**

| | |
|---|---|
| • Inquiries line | 1-800-635-7943 |
| • Fax line | 1-800-565-7757 |

## To access this product

This product, Catalogue no. 11F0019M, is available free in electronic format. To obtain a single issue, visit our website, www.statcan.gc.ca, and browse by "Key resource" > "Publications."

## Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "About us" > "The agency" > "Providing services to Canadians."

Published by authority of the Minister responsible for Statistics Canada



Cette publication est aussi disponible en français.

## Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

## Standard symbols

The following symbols are used in Statistics Canada publications:

| | |
|---|---|
| . | not available for any reference period |
| .. | not available for a specific reference period |
| ... | not applicable |
| 0 | true zero or a value rounded to zero |
| $0^s$ | value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded |
| $^p$ | preliminary |
| $^r$ | revised |
| x | suppressed to meet the confidentiality requirements of the *Statistics Act* |
| $^E$ | use with caution |
| F | too unreliable to be published |
| * | significantly different from reference category ($p < 0.05$) |

# The Effects of Cancer on Employment and Earnings of Cancer Survivors

by

Sung-Hee Jeon
Social Analysis and Modelling Division
Statistics Canada

11F0019M No. 362
ISSN 1205-9153
ISBN 978-1-100-25032-8

September 2014

**Analytical Studies Branch Research Paper Series**

The Analytical Studies Branch Research Paper Series provides for the circulation, on a pre-publication basis, of research conducted by Analytical Studies Branch staff, visiting fellows, and academic associates. The Analytical Studies Branch Research Paper Series is intended to stimulate discussion on a variety of topics, including labour, business firm dynamics, pensions, agriculture, mortality, language, immigration, and statistical computing and simulation. Readers of the series are encouraged to contact the authors with their comments and suggestions.

Papers in the series are distributed to research institutes, and specialty libraries. These papers can be accessed for free at www.statcan.gc.ca.

Publications Review Committee
Analytical Studies Branch, Statistics Canada
24th Floor, R.H. Coats Building
Ottawa, Ontario K1A 0T6

# Table of contents

## Abstract

The study examines the effects of cancer on the work status and annual earnings of cancer survivors who had a strong attachment to the labour market prior to their diagnosis. The comparison group consists of similar workers never diagnosed with cancer. The study is based on a Statistics Canada linkage file that combines microdata from the 1991 Census, the Canadian Cancer Registry, mortality records and personal income tax files. The study estimates changes in the magnitude of cancer effects during the first three years following the year of the diagnosis using a large sample of cancer survivors diagnosed at ages 25 to 61. The empirical strategy combines matching and regression models to deal with observed and unobserved differences between the cancer and comparison samples, and to improve causal inference. The results show moderate negative cancer effects on the work status and annual earnings. Overall findings suggest that, in the long run, cancer is more likely to affect survivors' work status than their earnings.



Keywords: cancer, employment, annual earnings

JEL classification: I1, J2, J3

## Executive summary


The study examines the effects of cancer on specific labour market outcomes of cancer survivors using the newly available linkage data combining Canadian 1991 Census microdata with administrative records from the Canadian Cancer Registry, Vital Statistics Registry and longitudinal personal income tax records. These unique data are used to estimate cancer effects on labour market outcomes of Canadian cancer survivors affected by all cancer types. Labour market outcomes of cancer survivors are compared to those of the comparison sample, drawn from the same linkage data files, consisting of individuals with similar characteristics but never diagnosed with cancer. The study specifically focuses on two labour market outcomes—employment status and annual earnings—of cancer survivors diagnosed at ages 25 to 61 who had a strong attachment to the labour market prior to the diagnosis. The study measures changes in the magnitude of cancer effects during three calendar years (*T*+1, *T*+2, and *T*+3) following the year of the diagnosis (*T*). The study also examines differences between cancer effects for subgroups with certain observable characteristics and the average cancer effect for the cancer sample as a whole.

The study results show that cancer survivors are less likely to work compared with individuals never diagnosed with cancer in all three post-diagnosis periods. The probability of working in *T*+1 is, on average, 3.0 percentage points lower in the group of cancer survivors relative to the comparison group. The difference in the probabilities increases to 3.7 percentage points in *T*+2 and 4.8 percentage points in *T*+3. The magnitude of the negative effect is greater for cancer types with low survival rates than for those with high survival rates. The probability of working in *T*+1 among cancer survivors with a low-survival-rate type of cancer is 11.0 percentage points lower than the probability of working for the comparison group. For those diagnosed with a high-survival-rate type of cancer, this probability is 2.0 percentage points lower than that of the comparison group.

The average annual earnings of cancer survivors are also lower than those of the comparison sample. The study found that cancer survivors earn, on average, about 12% less in *T*+1 than their counterparts never diagnosed with cancer when we accounted for transitions from employment to non-employment by including zero earners in the sample. When only those with non-zero annual earnings are considered, the estimated average earnings loss for cancer survivors in *T*+1 is 10% to 11%. In the second and third year after the year of the diagnosis, the earnings gap between cancer survivors and those in the comparison group narrows, particularly when cancer survivors continue to work after the cancer diagnosis. The effects of cancer on annual earnings are larger and longer lasting for survivors with low-survival-rate cancer, which is consistent with the results pertaining to employment status.

The study findings also suggest that, in the long run, the negative effect of cancer on the employment status for cancer survivors diagnosed at ages younger than the average age in the cancer sample, 48 years, is smaller than the average effect for the cancer sample as a whole. Conversely, the magnitude of the negative effect on the employment status is larger for cancer survivors with no high school diploma compared with the average effect in the sample. More generally, however, the patterns of post-cancer earnings losses among cancer survivors with different characteristics are mixed.

Observed labour market outcomes in the data are determined by the supply and demand sides of the labour market. However, the lack of information in the data makes it impossible to investigate the roles played by each side of the labour market in reducing the likelihood of employment of cancer survivors relative to those never diagnosed with cancer.

# 1 Introduction

About 1,638,910 new cancer cases were expected in the United States in 2012, and about 577,190 Americans were expected to die of cancer in that year (American Cancer Society 2012). In Canada, the Canadian Cancer Society (CCS) estimates there were about 186,400 new cancer cases and 75,700 cancer-related deaths in 2012 (CCS 2012). CCS data show that cancer incidence among the Canadian population steadily rose from 1983 to 2007—largely because of population growth and aging—even though age-standardized rates remained relatively stable during that period. In contrast, cancer mortality rates among men in Canada have slowly declined since the late 1980s and, among women, since the mid-1990s[1] (CCS 2012). Most member countries of the Organisation for Economic Co-operation and Development have seen declining cancer mortality rates since 1980 (OECD 2007). Earlier diagnoses, improvements in cancer treatment and better follow-up care have substantially increased the number of cancer survivors. The five-year relative survival rate[2] for all types of cancer in Canada increased from 56% in 1992–1994 to 62% in 2004–2006, based on 2004–2006 CCS estimates. Similarly, in the United States the five-year relative survival rate for all types of cancer diagnosed from 2001 to 2007 was 67%, up from 49% in 1975–1977, and the number of cancer survivors reached nearly 12 million in 2007 (American Cancer Society 2012; Centers for Disease Control and Prevention 2011).

This study examines the effects of cancer on labour market outcomes of cancer survivors. It uses linkage data combining Canadian 1991 Census microdata with administrative records from the Canadian Cancer Registry, the Vital Statistics Registry and longitudinal personal income tax records. These unique data are used to estimate cancer effects on labour market outcomes of Canadian cancer survivors affected by all types of cancer. Labour market outcomes of cancer survivors are compared with those of the comparison sample consisting of individuals never diagnosed with cancer and drawn from the same linkage data files. The study specifically focuses on the labour market outcomes of working-age cancer survivors with a strong attachment to the labour market prior to the diagnosis and the effects of cancer on the survivors' labour market outcomes approximately one to four years from the time of the cancer diagnosis. The study measures changes in the magnitude of negative labour market effects associated with cancer during the three years after the calendar year of the diagnosis. Differences between cancer effects for subgroups with certain observable characteristics and the average cancer effect for the cancer sample as whole are also examined.

As cancer survivor rates rise, researchers are increasingly trying to understand the effect of cancer on economic outcomes of cancer survivors. Breast cancer has so far received the most attention in this literature, partly because it is the most common type of cancer among individuals of working age and partly because individuals diagnosed with breast cancer have higher survival rates compared to those with other types of cancer. Among recent studies, Bradley et al. (2005) find that breast cancer survivors aged 30 to 64 are less likely to work shortly after the diagnosis (within the next six months) compared with other women. However, when Bradley et al. consider a longer post-diagnosis period, 12 to 18 months, they find little evidence of any negative effect (Bradley et al. 2007). Moran et al. (2011), who study the long-term (two to six years after the diagnosis) effects of all cancer types on employment of workers aged 28 to 54, find that cancer survivors have lower employment rates than other similarly aged adults.

Much of the literature on economic outcomes of cancer survivors is fairly recent, and the evidence presented so far is mixed, particularly regarding long-term effects. Scarcity of population-based data that contain both cancer diagnosis and labour market information has

1. Breast cancer mortality rates among women have declined since the mid-1980s (Canadian Cancer Society 2012).
2. The five-year relative survival rate is the probability of surviving for five years after a cancer diagnosis compared to the probability of surviving for five years in the general population of the same age and sex.

been a major hindrance for researchers. Common population-based household surveys containing labour market information generally do not contain cancer information: even if cancer survivors can be identified in such surveys—generally older workers—the sample size is very small (Bradley et al. 2002). Other data sources may contain labour market information by surveying a sample of cancer survivors identified based on cancer patients records, in which case labour market information for a comparison sample have to be obtained from other data sources (Bradley et al. 2005; Moran et al. 2011). However, in these cases, data are collected from a targeted local area and may not be representative at the national level. Another common problem is the lack of longitudinal labour market outcome data—particularly earnings data—for any extended period of time before and after a cancer diagnosis. Previous studies that examined long-term effects of cancer on employment outcomes used retrospective questions to construct pre-diagnosis labour market histories (e.g., Moran et al. 2011), in which case the results are likely to suffer from substantial measurement error.

This study considers two labour market outcomes: work status (working or not working); and total annual earnings. A particular strength of the study is that these labour market outcome variables are based on highly accurate individual-level data from annual personal income tax files over pre- and post-cancer periods. Yet, unlike most survey-based studies, this study cannot explicitly examine the effect of cancer on the labour supply of cancer survivors—their labour force participation or hours of work—because these variables, routinely available in survey data, are not available in administrative data. Conceptually, however, there are at least two important ways in which a cancer diagnosis can affect individuals' labour supply decisions and, consequently, their labour market outcomes. First, cancer alters the allocation of individuals' time: they must spend some of it on cancer treatment and recovery. Accordingly, cancer survivors may have less time that can be allocated for work, so they may reduce their working hours or stop working altogether. However, it is likely to result in a temporary decline in individuals' labour supply and, consequently, a decrease in their employment or earnings would not last long. Second, not only can cancer and cancer treatment physically constrain the amount of time cancer survivors are able to work during a day; like other health shocks, cancer also can shift individuals' preferences away from work toward leisure. In this case, the reduction in labour supply would last longer and may become permanent. This would result in a steady decline in employment or earnings observed in longitudinal data.[3]

Following recommendations in Ho et al. (2007), this study combines matching with parametric models to improve causal inferences for the estimates based on observational data. The data are preprocessed using the Coarsened Exact Matching (CEM) method and reweight the comparison sample using CEM matching weights to make the comparison sample as similar as possible to the cancer sample (Iacus et al. 2012). The weighted regression accounts for potential confounding effects of observable pre-cancer characteristics such as age and education on cancer incidence and labour market outcomes of cancer survivors. Also, the study attempts to control for the potential correlation among unobservable characteristics, cancer incidence and labour market outcomes by combining matching with the regression-adjusted 'difference-in-differences' (DID) method (Heckman et al. 1997, 1998).

The results show that the employment rate among all cancer survivors was lower than the employment rate among individuals never diagnosed with cancer in all three post-diagnosis periods. However, the estimated magnitude of the difference in the employment rates is relatively moderate (3 percentage points in the first year after cancer diagnosis). The magnitude of the negative effect is greater for cancer types with low survival rates than for those with high

3. Observed labour market outcomes are determined by the supply and demand sides of the labour market. The data contain no information about employers' responses to employees' cancer diagnoses. The situation becomes even more complicated if the cancer or cancer treatment becomes associated with a prolonged severe disability limiting cancer survivors' work or job performance. However, cancer-related disabilities are not observed in the data.

survival rates. The average annual earnings of cancer survivors are also moderately lower than the average earnings of the comparison sample (12.1% at the first year after cancer diagnosis). However, the earnings gap between cancer survivors and their counterparts in the comparison sample narrows, particularly when survivors continue to work after the cancer diagnosis. Overall, the findings suggest that in the long run, cancer is more likely to affect survivors' work status than their earnings.

The plan of the paper is as follows. Section 2 describes the empirical strategy of identifying the causal impact of cancer on labour market outcomes. Section 3 describes all data sources, and explains the selection rule for the cancer and comparison samples. Descriptive statistics for the cancer and comparison samples before and after matching are also provided in Section 3. Section 4 presents results. Section 5 concludes.

# 2 Empirical strategy and model

## 2.1 Observable confounding factors affecting cancer and labour market outcomes

A first look at the data reveals that cancer does not affect individuals in the sample randomly. As well, the incidence of cancer is correlated with some of the observable characteristics: one indication is that the average characteristics of the cancer sample are different from those of the comparison sample. Not accounting for these factors would significantly bias the results. (Detailed summary statistics of both samples will be presented in the next section.) The first concern is that factors increasing the chances of having cancer also have a direct impact on labour market outcomes. Aging, for example, increases cancer incidence and at the same time decreases the probability of working. To account for the confounding influence of pre-cancer conditions on cancer incidence and outcome variables, it is important to make characteristics of the comparison sample more similar to those of the cancer sample.

The study applies the CEM method to match the cancer sample with the comparison sample based on the pre-cancer period observable characteristics (Iacus et al. 2012). The CEM is a multidimensional exact matching algorithm applied to cells generated by dividing continuous variables into discrete intervals or by regrouping categorical variables into fewer coarsened categories.[4] The CEM algorithm creates a set of strata with the same coarsened values of matching variables; it also restricts the matched data to areas of common empirical support by pruning unmatched observations from both treated and control samples. For each stratum, the CEM returns weights $(n_t / n_c * N_c / N_t)$[5] that can be used to reweight observations in the matched comparison sample and balance the empirical distributions of the matching variables between the cancer and comparison samples.

Ho et al. (2007) demonstrate that preprocessing raw data using matching procedures turns parametric models into a much more reliable tool of the empirical analysis of causal effects. In

4. The CEM method dominates other matching methods (e.g., propensity score matching) commonly used to reduce imbalance related to the first and higher moments, nonlinearities, interactions, and other multidimensional distributional difference between the treated and control groups. See Iacus et al. (2012, 2011) for a detailed discussion of CEM properties and comparison with other matching methods.

5. Weights assigned to the matched comparison sample will be equal to the ratio of the cancer sample size $(n_t)$ to the comparison sample size $(n_c)$ in each stratum multiplied by the ratio of the total size of the comparison sample $(N_c)$ to the total size of the cancer sample $(N_t)$. The weights for the matched cancer sample are unities. The weights for unmatched records are zeros.

particular, estimates of causal effects are less sensitive to the choice of a model specification.[6] One of the proven properties of the CEM is that it reduces the degree of model dependence (Iacus et al. 2012).[7] To make causal inference, matching weights are used in the regression analysis of labour market outcomes. As well, the regression method controls for the remaining imbalance since, depending on coarsening, some imbalance can remain in the matched data. The estimates of the cancer effect in the model measure the average effects of cancer on the outcome variables for cancer survivors.[8]

## 2.2 Unobservable confounding factors affecting cancer and labour market outcomes

Another concern is the correlation between cancer incidence and individuals' unobservable characteristics—this is also known as 'selection-on-unobservables' in the evaluation literature. Like observed differences, unobserved differences between the cancer and comparison samples may also result in different labour market outcomes for these two groups. For instance, developing cancer may be correlated with unhealthy lifestyle unobserved in the data (e.g., smoking or poor diet). This, in turn, may depend on particular unobserved individual characteristics related to personal motivation. Individuals with such characteristics would have less chance of developing cancer (e.g., by refraining from smoking) and, at the same time, have greater likelihood of earning higher income. Even if the confounding effect of the unobserved characteristics in this example is not particularly strong, not controlling for the average difference in unobservable characteristics of the cancer and comparison groups would lead to overestimating the negative effect of cancer on labour market outcomes.[9]

Heckman et al. (1997, 1998) proposed combining DID and matching approaches to eliminate time-invariant unobservable differences between two groups. This approach, however, requires pre- and post-diagnosis data, which are now included in the newly available linkage data used in this study. Adopting the Heckman et al. (1997, 1998) strategy, the difference in individuals' earnings before and after the cancer diagnosis is used as a dependent variable regressed on the cancer variable using CEM matching weights.[10]

## 2.3 Model specification

In the statistical model, $T$ is the year of cancer diagnosis in the cancer sample and the matched year in the comparison sample. Calendar years before and after the year of cancer diagnosis (or years matched to them in the comparison sample) are $T-j$ and $T+j$, where $j=1,...,J$. Labour market outcomes at $T$+1, $T$+2 and $T$+3 are modelled as a function of a cancer indicator and individual as well as work characteristics

$$\textit{Labour market outcome} = \mathrm{f}(\textit{cancer, individual characteristics, work characteristics}) + \varepsilon.$$

Individual characteristics include age, age squared, marital status (couple/single), an indicator of having a long-term disability in 1991 and province of residence. The variables used for

6. Ho et al. (2007) recommend that researchers use parametric models to adjust the matched sample. They characterize the advantage of two-step procedures as doubly robust in the sense that, under weak conditions, if either the matching or the parametric model is correct—but not necessarily both—causal estimates will be consistent.
7. Model dependence is defined by how much the predicted value of the outcome variable varies as a function of the statistical model for a given set of explanatory variables (Ho et al, 2007). One of the key reasons for matching is to eliminate model dependence; however, it has never been proven mathematically for any previous matching methods commonly used in the various analyses. For a detailed discussion, see Iacus et al. (2011).
8. In evaluation literature, this is known as 'the sample average treatment effect on the treated'.
9. Moran et al. (2011) express a similar concern in their study of cancer effects on employment.
10. By differencing individual pre- and post-cancer earnings, it is also possible to account—at least partly—for unobserved individual fixed effects that are not necessarily correlated with cancer (e.g., individual's work ethics).

matching—sex, education, visible minority status—are also included in the model specification.[11] Work characteristics are indicators for having non-zero self-employment income and union (professional association) membership. Quintile dummies for total earnings at $T-1$ and year dummies are also included in the model.

# 3 Data

The 1991 Census–Longitudinal Worker File (LWF) is a unique dataset that combines data from four sources: Canada's 1991 Census of Population, the Canadian Mortality Database (CMDB), the Canadian Cancer Database (CCDB) and the LWF.

The CMDB contains individual death records from 1950 onward. Provincial and territorial vital statistics offices provide these records annually to Statistics Canada for national-level analysis.

The CCDB is a databank combining two cancer data sources: the Canadian Cancer Registry (CCR) and the National Cancer Incidence Reporting System (NCIRS). The former is a person-oriented tumor database, which includes clinical and demographic information about Canadian residents with cancer since 1992 (Statistics Canada 2008). The latter is a historical tumor-oriented database containing cancer cases diagnosed as far back as 1969 (Carpenter et al. 2008). Individual cancer records from the CCR are used in the analysis; historical information from NCIRS is used to verify that individuals in the CCR had no prior cancer history.

The LWF represents 10% of the random sample of Canadians who either filed a personal income tax form (Form T1 General, *Income Tax and Benefit Return*) or received a statement of remuneration (Form T4, *Statement of Remuneration Paid (slip)*) from their employers in each year from 1983 onward. Once individuals are selected into the LWF, they are followed regardless of their employment status for as long as they file tax returns (Form T1 General) or their incomes are reported to the Canada Revenue Agency (CRA) by their employers. The current version of the LWF contains information on wages, salaries and net self-employment income as well as firm-level information.[12] Wages and salaries are obtained from T4 slips issued by employers. Net self-employment income and basic personal information (marital status, province of residence, etc.) are obtained from the personal income tax files (T1).[13]

Statistics Canada's Health Analysis Division initially linked selected personal information from CMDB and CCDB to the individual records of individuals aged 25 and older in the 1991 Census file.[14] This initial data linkage is called '1991 Canadian Census Cohort: Mortality and Cancer Follow-up.' Individuals' death records up to 2006 and individuals' cancer records up to 2003 were obtained from both data banks, CMDB and CCDB.[15] Subsequently, the LWF records were linked to the 1991 Canadian Census Cohort to provide the crucial income component.

The 1991 Census–LWF data sample contains 263,674 individual records corresponding to about 1.4% of the Canadian population aged 25 and over in 1991. Approximately 58.8% of the 1991 Census–LWF cohort was observed in all 28 years of the LWF, from 1983 to 2010; the

11. This does not change the cancer effect (the coefficient on the cancer variable) since with reweighting, the cancer is mean-independent conditioning on these matching variables.
12. Net self-employment income is income from unincorporated businesses; income from incorporated businesses is reported as wages and salaries.
13. Filing rates in Canada are very high primarily because tax benefits, such as refundable Goods and Service Tax (GST) credits, give incentives to file tax returns even to individuals with low or no income.
14. At that stage, the linkage was made using Statistics Canada probabilistic record linkage methods. See Wilkins et al. (2008) for detailed information on the data linkage. The second stage linkage of the LWF to the 1991 Census cohort was made through a deterministic record linkage process based on Social Insurance Number (SIN).
15. Further data development extending the data on the 1991 Census cohort to more recent years is currently in progress. Detailed information on the 1991 Census and both databases is available from the Statistics Canada website (www.statcan.gc.ca).

average number of years an individual is present in the sample is 24.8. Tax-filing rates were slightly lower in the 1980s compared with the more recent decades; in the years 1990 to 2010, 66.9% of 263,638 individuals are observed in all 21 years, and the average number of years an individual is observed in the sample is 18.5.

## 3.1 Cancer sample

Initially, individuals diagnosed with cancer for the first time from 1992 to 2000 are selected from the 1991 Census–LWF cohort. As this study focuses on the labour market outcomes of the cancer survivors, the sample is restricted to individuals who were under 62 when they were diagnosed with cancer. In the 1991 Census–LWF data, only cancer sites are available, and no information about the severity and stage of cancer is available in the data. First, all tumors are grouped according to 26 of the most common cancer sites based on the Surveillance, Epidemiology, and End Results (SEER)[16] grouping for ICD-9 (International Statistical Classification of Diseases, 9th Edition) and ICD-O-2/3 (International Classification of Diseases for Oncology, Second and Third Edition) codes available in the data. For those who had multiple tumor records in any year, the record with a malignant and primary site tumor that has the lowest relative survival rate according to previous studies on cancer survival rates (Ellison et al. 2011) is selected. Then, 26 sites are grouped into three categories—high, middle and low—based on the five-year relative survival ratio (hereafter referred to as the 'cancer survival category') for each cancer site (Ellison et al. 2011). These cancer categories are used as a proxy for the average severity of cancer to determine the impact of cancer severity on labour market outcomes of cancer survivors.[17]

The 1991 Census–LWF data contains 5,185 individuals (2,120 males and 3,065 females) diagnosed with cancer for the first time from 1992 to 2000 who were aged 61 or younger in the year of diagnosis.[18] The numbers of cases for each of the 26 cancer sites are shown in Appendix Table A.1. Among all 5,185 cancer cases, about 8% (412) are tumors other than the 26 common sites. The most common cancer sites are breast (21.5%), lung and bronchus (10%), cervix uteri (7.9%), prostate (5.9%) and colon (5.8%). Broken down by cancer-survival categories, 2,023 (39%) cases are grouped as high, 1,659 (32%) as middle, and 1,091 (21%) as low. Also, 2,020 records (39% of the total sample) are matched to death records from CMDB mortality data (the last available year is 2006). By cancer-survival categories, 19.1% of the high, 34.3% of the middle, and 81.5% of the low-survival groups are matched to death records.

For the analysis, 3,716 individuals who survived for at least three years after their cancer diagnosis are kept in the sample, and 77 individuals who had multiple new cancer diagnoses within those three years are excluded from the sample. Finally, based on their labour market attachment prior to the cancer diagnosis, the cancer sample is restricted to individuals who work in the year of cancer diagnosis ($T$) and two previous years ($T-1$ and $T-2$) since the focus of the study is on the effects of cancer on labour market outcomes for those with a strong labour market attachment prior to the cancer diagnosis.[19] More details on the labour market outcome measures will be presented in Subsection 3.3. By imposing this restriction, 1,042 individuals are further excluded from the cancer sample. To summarize, the cancer sample consists of 2,597 cancer survivors diagnosed with cancer for the first time from 1992 to 2000 and aged 61 or younger at the time of the diagnosis. These individuals worked in the year of the diagnosis and

16. http://seer.cancer.gov.

17. Differences in the average survival rates are not necessarily equivalent to differences in morbidity across different cancer types. However, no suitable morbidity index that could be used as a more reliable measure of severity for all cancer types is available.

18. Among them, 219 people had multiple cancer diagnoses between 1993 and 2003.

19. Two alternative samples are constructed to determine the robustness of the results to different restrictions on the individual's work status prior to a cancer diagnosis. One sample contains individuals working at $T$ and $T-1$ and the other sample contains individuals working at $T$, $T-1$, $T-2$ and $T-3$. All findings are very robust to different sample restrictions. See also Footnote 25 for more details.

two previous years, survived for at least three years after the diagnosis and had no subsequent diagnoses during those three years.

Table 1 shows the types of cancer in the cancer sample. Since the cancer sample is restricted to people who were diagnosed with cancer and survived for more than three years, it contains proportionally fewer people in the low-survival category than shown in Appendix A. One-half of the cancer sample is categorized as a high-survival group. General characteristics of the cancer sample are presented in the second column of Table 2 in the next subsection.

**Table 1**
**Cancer types (sites) in the cancer sample by survival category**

| | Share of cancer sample | |
|---|---|---|
| | **Type** | **Survival** |
| | percent | |
| **High-survival category** | ... | 50.10 |
| Thyroid | 3.93 | ... |
| Prostate | 7.74 | ... |
| Testis | 1.85 | ... |
| Skin melanoma | 6.93 | ... |
| Breast | 25.68 | ... |
| Corpus uteri | 3.20 | ... |
| Hodgkin lymphoma | 0.77 | ... |
| **Middle-survival category** | ... | 34.35 |
| Chronic lymphocytic leukemia | 0.85 | ... |
| Cervix uteri | 10.63 | ... |
| Bladder (including in situ) | 3.43 | ... |
| Kidney and renal pelvis | 2.89 | ... |
| Soft tissue | 0.89 | ... |
| Larynx | 0.85 | ... |
| Rectum | 3.50 | ... |
| Colon | 4.74 | ... |
| Non-Hodgkin lymphoma | 4.35 | ... |
| Oral (buccal cavity and pharynx) | 2.23 | ... |
| **Low-survival category** | ... | 8.09 |
| Ovary | 1.96 | ... |
| Multiple myeloma | 0.62 | ... |
| Leukemia (excluding chronic lymphocytic leukemia) | 0.81 | ... |
| Stomach | 0.35 | ... |
| Brain | 1.19 | ... |
| Liver | x | ... |
| Lung and bronchus | 2.66 | ... |
| Esophagus | x | ... |
| Pancreas | 0.19 | ... |
| Others | 7.47 | ... |
| | number | |
| Total number of cancer survivors | 2597 | ... |

... not applicable

x suppressed to meet the confidentiality requirements of the *Statistics Act*

**Note:** Author's calculations.

**Source:** Statistics Canada, 1991 Census–Longitudinal Worker File.

## 3.2 Comparison sample

This section describes the selection of the comparison sample whose labour market outcomes are compared with those of the cancer sample. The general selection criteria used to select the cancer sample are first applied: individuals are pre-selected based on their cancer records, year, age, death records, and work history from the same 1991 Census–LWF data. Individuals of the 192,537 who have no cancer records in all available years in either the CCR or NCIRS (historical tumor records) are initially chosen from the data. Similarly to the cancer sample, the comparison sample consists of individuals aged 61 or younger in each year from 1992 to 2000. Then, in each year, individuals who lived for at least three years following $T$ (that excludes 1,053 individuals) and worked in year $T$ and two previous years ($T-1$ and $T-2$) (further excluding 22,576 individuals) are kept in the sample. At this point, 168,908 individuals are pre-selected from the 1991 Census–LWF data and the pooled number of observations over nine years from 1992 to 2000 is 1,228,551.

On the next step, CEM is applied to match this pre-selected sample with the cancer sample to obtain the matched comparison sample. Matching variables include age in year $T$, gender, level of education, visible minority status, year $T$ and province of residence in year $T-1$. Age is coarsened into five-year intervals. The visible minority status is coarsened into three categories: non-minority, Asian and other minority. The Northwest Territories, Nunavut and Yukon are merged into a single category (Northern Territories). CEM automatically restricts the cancer and comparison samples to areas of common empirical support by pruning unmatched observations from both samples. Four among 2,597 individuals in the cancer group cannot be matched to anyone in the comparison group. The matched comparison group consists of 624,835 observations (pooled year data for 142,196 individuals).[20] Table 2 shows the summary statistics for matching variables for the cancer and comparison group before and after matching. The summary statistics for other variables, not used in the matching procedure but used in the regression analysis, are provided in Appendix Table A.2. The last column of Table 2 presents the differences in the variables' means (proportions) between the cancer and comparison samples before matching and t-tests for the differences between these two samples.[21]

20. Since the same individuals may appear in the comparison sample for more than one matching year, standard errors in all the regression models are clustered on individual identifiers.

21. In general, previous studies using matching methods report normalized before-and-after mean differences of covariates rather than the t-statistics (or t-test results). This is done to show the improvement in the mean balance between the two samples because—unlike t-statistics—the normalized mean differences do not depend on the unmatched or matched sample size (Imbens and Wooldridge 2009). In this study, however, as a result of CEM, there are no remaining mean differences between the cancer and the matched comparison sample (as Table 2 shows). Table 2 presents t-test results only to show that, for the matching variables, mean differences between the two samples had been statically significant before the matching procedure was applied.

## Table 2
**Summary statistics for the cancer and comparison samples before and after matching and pruning**

| | Cancer sample | Comparison sample: Unmatched | Comparison sample: Matched and reweighted | Unmatched difference |
|---|---|---|---|---|
| | | percent | | percentage points |
| Age (mean) at *T* | 48.12 | 43.05 | 47.97 | 5.07 *** |
| **Coarsened age at *T*** | | | | |
| 25 to 29 | 0.92 | 3.08 | 0.93 | -2.16 *** |
| 30 to 34 | 6.62 | 13.92 | 6.63 | -7.30 *** |
| 35 to 39 | 11.05 | 20.83 | 11.07 | -9.78 *** |
| 40 to 44 | 14.13 | 20.29 | 14.15 | -6.16 *** |
| 45 to 49 | 18.33 | 17.72 | 18.36 | 0.61 |
| 50 to 54 | 21.72 | 13.44 | 21.75 | 8.28 *** |
| 55 to 59 | 20.64 | 8.48 | 20.63 | 12.16 *** |
| 60 to 61 | 6.59 | 2.24 | 6.48 | 4.34 *** |
| Female | 62.84 | 46.02 | 62.86 | 16.82 *** |
| **Highest level of schooling** | | | | |
| No high school | 25.45 | 22.21 | 25.49 | 3.25 *** |
| High school with/without trades | 39.24 | 40.86 | 39.26 | -1.62 † |
| Postsecondary non-university | 19.21 | 19.23 | 19.24 | -0.02 |
| University degree | 16.10 | 17.70 | 16.01 | -1.61 * |
| **Visible minority** | | | | |
| Non-minority | 93.96 | 92.03 | 94.06 | 1.93 *** |
| Asian | 3.97 | 5.26 | 3.93 | -1.30 ** |
| Other | 2.08 | 2.71 | 2.01 | -0.63 * |
| **Province/region at *T*-1** | | | | |
| Newfoundland | 1.54 | 1.82 | 1.58 | -0.28 |
| Prince of Edward Island | 0.77 | 0.51 | 0.74 | 0.26 † |
| Nova Scotia | 3.66 | 3.12 | 3.58 | 0.54 |
| New Brunswick | 2.97 | 2.53 | 2.85 | 0.43 |
| Quebec | 23.72 | 25.29 | 23.81 | -1.57 † |
| Ontario | 32.19 | 36.18 | 32.35 | -3.99 *** |
| Manitoba | 4.81 | 4.06 | 4.85 | 0.76 † |
| Saskatchewan | 3.58 | 3.60 | 3.54 | -0.02 |
| Alberta | 12.01 | 9.98 | 12.09 | 2.03 ** |
| British Columbia | 13.98 | 11.94 | 13.83 | 2.04 ** |
| Northern Territories | 0.77 | 0.93 | 0.77 | -0.16 |
| Missing | 0.00 | 0.04 | 0.02 | -0.04 |
| **Year at *T*** | | | | |
| 1992 | 9.90 | 12.18 | 9.91 | -2.29 ** |
| 1993 | 10.44 | 11.84 | 10.41 | -1.40 * |
| 1994 | 10.71 | 11.50 | 10.68 | -0.79 |
| 1995 | 10.94 | 11.27 | 10.91 | -0.34 |
| 1996 | 11.55 | 11.05 | 11.57 | 0.50 |
| 1997 | 11.13 | 10.84 | 11.15 | 0.29 |
| 1998 | 11.36 | 10.58 | 11.38 | 0.78 |
| 1999 | 12.17 | 10.43 | 12.15 | 1.73 ** |
| 2000 | 11.82 | 10.30 | 11.84 | 1.52 * |
| | | number | | |
| *n* | 2,597 | 1,228,551 | 624,835 | ... |

... not applicable

*** significantly different from reference category ($p<0.001$)

** significantly different from reference category ($p<0.01$)

* significantly different from reference category ($p<0.05$)

† significantly different from reference category ($p<0.10$)

**Note:** The last column displays mean differences (proportion) between cancer and comparison samples before matching; standard symbols indicate significance of t-test differences between these two groups. *T*s are matching years which are years of cancer diagnosis for the cancer sample. *n*: number of observations. *T*-1: *T* minus 1. Author's calculations.

**Source:** Statistics Canada, 1991 Census–Longitudinal Worker File.

The average age in the cancer sample is higher than that of the pre-matched comparison sample. The age distribution shows that cancer incidence is positively correlated with age. The cancer sample also has proportionally more individuals with less than a high school level of education than the comparison sample. Recent literature finds evidence of causal influence of higher education attainment on health and healthy lifestyle (Jones et al. 2011). The cancer sample has proportionally more women than the pre-matched comparison sample. Breast cancer, the most common type of cancer among women, is also the most prevalent cancer type in the study sample. Among men, prostate cancer is the most common type of cancer, but it usually develops later in life at ages excluded from the study sample. The data are also consistent with the fact that prevalence of cancer is generally lower among Asians than among non-minorities.

Shown in the fourth column, the reweighted (post-matching) distributions are almost identical in the cancer and the matched comparison samples. Matching and reweighting procedures enables us to regard cancer as randomly assigned "treatment" conditional on the matched observable (pre-cancer) sample characteristics.

Summary statistics for earnings in the two years ($T$-1 and $T$-2) prior to cancer diagnoses in the cancer sample are presented in Appendix Table A.2 along with the corresponding earnings in the matched comparison sample. As a result of matching, both the average earnings and the earnings distributions in the cancer sample in the years before the cancer diagnosis are virtually identical to those in the matched comparison sample. This observation has two implications. First, total annual earnings can be accurately predicted in the study samples based on variables used for matching. Second, in the absence of cancer diagnosis, the average earnings of the cancer sample would follow the same common trend as those of the matched comparison sample. This observation is particularly important, as it is consistent with the main assumption of the DID approach used in the empirical analysis.[22]

## 3.3 Measures of labour market outcome

The study considers two labour market outcome variables: the first is a binary annual working status variable; the second, a continuous variable for total annual earnings (in 2010 dollars). Individuals' total annual earnings are defined as the sum of all their wages and salaries received in a given year plus net self-employment income in that year. If individuals had non-zero earnings—either positive or negative—they are considered to have worked in that year. In the data, the frequencies of outcome variables are annual. There is no information on how many months or weeks each cancer survivor worked before and/or after the cancer diagnosis in the year he or she was diagnosed. To ensure that the person had not stopped working before being diagnosed with cancer, the cancer sample is restricted to individuals with non-zero earnings in the year of the cancer diagnosis ($T$) as well as two previous years ($T$-1 and $T$-2). The study examines the effect of cancer on labour market outcomes in the years after the diagnosis by comparing labour market outcomes of the cancer and comparison samples for three consecutive periods: $T$+1, $T$+2 and $T$+3.

Note that $T$ corresponds to the calendar year, not the actual time of the diagnosis. Accordingly, $T$+$j$ corresponds to the calendar years following the year of diagnosis, not the time intervals that elapsed from the moment of the actual diagnosis. For instance, suppose that a cancer survivor had zero earnings in year $T$+1 (the year after the year of diagnosis). It is known that this person did not work for at least 12 months, but that period may be as long as 24 months from the time of the diagnosis if the person was diagnosed at the beginning of year $T$. However, if a cancer survivor had positive earnings in year $T$+1, these earnings reflect total earnings for a period of 12 to 24 months from the cancer diagnosis. Therefore, the actual working time from the day of

22. The underlying assumption for causal interpretation is that, in the absence of cancer diagnosis, the earnings of the cancer sample follow the same common trend as those of the comparison sample.

the cancer diagnosis to the end of $T$+1 depends on the time of the cancer diagnosis in year $T$. Consequently, the data are not suitable for analyzing the short-term (e.g., less than a year) effects of cancer diagnosis on labour market outcomes; thus, only longer-term labour market outcomes are considered. Appendix Table A.3 shows unadjusted means of employment and earnings in $T$+1, $T$+2 and $T$+3 by cancer status and cancer-survival category.[23] Unadjusted mean differences in employment and earnings between the cancer sample and the matched comparison sample are -4.36% and $-9,022 at $T$+1.

# 4 Results

## 4.1 The effect of cancer on employment

The left panel of Table 3 shows estimates of cancer effects on the probability of working at $T$+1, $T$+2 and $T$+3. Because all individuals in the sample worked at $T$, $T$-1, and $T$-2, estimating the probability of working in each period is equivalent to estimating the probability of transition from employment to non-employment between $T$ and $T$+1, $T$ and $T$+2, and $T$ and $T$+3, respectively.[24] Here, the effect of cancer on employment at the end of each time interval combines various aspects of all labour market transitions that took place between $T$ and the final period. Some individuals, for instance, may temporarily stop working after the cancer diagnosis because of cancer treatment and post-treatment recovery but return to work in the subsequent periods. Others may not leave the labour market immediately after the cancer diagnosis but will leave in the subsequent periods. Some individuals may leave the labour market permanently following the cancer diagnosis. The differences between the survivor and comparison groups in the probability of working at the end of each period reflect the cumulative effect of cancer on employment from $T$ to that period.

The upper-left panel shows that the probability of working in $T$+1 is, on average, 3.0 percentage points lower in the cancer group relative to the comparison group. The difference in the probabilities increases to 3.7 in $T$+2 and 4.8 percentage points in $T$+3.

The right panel of Table 3 presents the effects of cancer on the probability of working conditional on continuously working after the cancer diagnosis until $T$+1, $T$+2 or $T$+3. This is equivalent to estimating the probability of exit from employment between two consecutive periods (a more conventional exit rate). The results in the upper-right panel suggest that even if cancer does not interrupt individuals' attachment to the labour market soon after the diagnosis, it may affect their exit from the labour market in the longer term. The exit rate decreases with more years since the cancer diagnosis, but it is still 1.3 percentage points higher at $T$+3 than the average exit rate for people never diagnosed with cancer, and the difference is statistically significant.

In the bottom panel, instead of a binary cancer variable in the model, three cancer-survival categories—high, middle and low—are included in the estimation model. Among these three categories, the negative effect of cancer on employment is the largest for the cancer survivors in the low-survival category in all three periods compared with people never diagnosed with cancer. In the bottom panel, the probability of working at $T$+1 for survivors in the low-survival group is 11 percentage points lower than the probability of working in the comparison group. In addition, the bottom right panel shows that, for those in the low-survival category, the exit rates conditional on working in all previous years also remain high and statistically significant for all three post-diagnosis periods. However, in the high-survival category, for those who continue to

23. Although tax filing rates in Canada are very high, individuals' data in $T$+1, $T$+2 and $T$+3 can be missing, so the panel is unbalanced. Immigration/emigration, late filing or non-eligibility for benefits can be possible reasons for missing tax years.

24. Interpreting estimates as exit probabilities would require simply reversing the signs.

work after the cancer diagnosis, the estimated exit rate at *T*+3 is no longer significantly different from the average in the comparison group.

**Table 3**
**The effect of cancer on employment**

| | Work in each period: Pr($work_{T+j}$=1) | | | Work continually up to each period: Pr($work_{T+j}$=1 \| $work_{T+j-1}$=1, $work_{T+j-2}$=1) | | |
|---|---|---|---|---|---|---|
| | *T*+1 | *T*+2 | *T*+3 | *T*+1 | *T*+2 | *T*+3 |
| **Cancer** | | | | | | |
| Coefficient | -0.030 *** | -0.037 *** | -0.048 *** | -0.030 *** | -0.018 *** | -0.013 ** |
| Standard error | 0.005 | 0.006 | 0.007 | 0.005 | 0.005 | 0.005 |
| *n* (weighted sample) | 588,593 | 586,698 | 584,994 | 588,593 | 567,922 | 547,901 |
| **Survival category** | | | | | | |
| **High** | | | | | | |
| Coefficient | -0.021 ** | -0.036 *** | -0.039 *** | -0.021 ** | -0.025 *** | -0.008 |
| Standard error | 0.007 | 0.008 | 0.009 | 0.007 | 0.007 | 0.007 |
| **Middle** | | | | | | |
| Coefficient | -0.034 *** | -0.025 * | -0.024 * | -0.034 *** | -0.001 | -0.001 |
| Standard error | 0.009 | 0.010 | 0.011 | 0.009 | 0.007 | 0.007 |
| **Low** | | | | | | |
| Coefficient | -0.107 *** | -0.129 *** | -0.206 *** | -0.107 *** | -0.080 ** | -0.081 ** |
| Standard error | 0.025 | 0.028 | 0.031 | 0.025 | 0.025 | 0.027 |
| *n* (weighted sample) | 588,411 | 586,514 | 584,815 | 588,411 | 567,746 | 547,733 |

*** significantly different from reference category ($p<0.001$)

** significantly different from reference category ($p<0.01$)

* significantly different from reference category ($p<0.05$)

**Note:** Results are average partial effects from the probit model with Coarsened Exact Matching (CEM) weights. All regressions are conditional on the full set of control variables and are listed in Subsection 2.3. In the left panel, the self-employment and union membership variables correspond to period *T*. In the right panel, lagged values of the self-employment and union membership variables are included as controls. *n*: number of observations; *T*: year of cancer diagnosis for the cancer sample. Author's calculations.

**Source:** Statistics Canada, 1991 Census–Longitudinal Worker File.

## 4.2 The effect of cancer on earnings

Next, the effect of cancer on total annual earnings is estimated using a weighted linear regression. The left panel of Table 4 shows estimates of cancer effect in each period not conditional on working. Zero annual earnings are included in this analysis, so the results presented in the left panel include earnings losses resulting from transitions from employment before cancer diagnosis to non-employment after the diagnosis. Estimates in the right panel of Table 4 are conditional on working in each period (zero earnings excluded). These results are close to the cancer effect along the intensive margin of labour market outcomes and can be compared to the estimated effects along the extensive margin shown in Table 3.

Cancer survivors earn, on average, $5,079, or 12.1%, less at *T*+1 than their counterparts never diagnosed with cancer; however, the negative effect of cancer is smaller at *T*+2, 9.7%, and smaller still at *T*+3, 9.3% (upper-left panel in Table 4). When the effect of cancer is estimated conditional on working, workers diagnosed with cancer at *T* earn $4,675 (10.6%) less at *T*+1 than do their counterparts from the comparison sample. The negative effect of cancer is smaller at *T*+2, 6.6%, and *T*+3, 5.4%, than at *T*+1.Comparing these results with the effects of cancer on the probability of working shown in Table 3, the impact of cancer on work status appears to be more persistent than its impact on annual earnings in post-diagnosis periods.

The bottom panel of Table 4 shows results for the model in which the cancer variable is grouped into three survival categories. In both right and left panels, the earnings decline, relative to the comparison sample, is greater for survivors in the low-survival category than for those in the high-survival category. The effects of cancer are larger and longer lasting for the low-survival category than for other categories, which is consistent with the results for the employment status shown in Table 3. As time goes by, the average annual earnings of cancer survivors remaining in the labour market seem to converge to the average annual earnings of workers with similar characteristics never diagnosed with cancer, particularly for those in the high-survival category.

**Table 4**

**The effect of cancer on annual earnings (in dollars)**

| | Earnings in each period not conditional on working in that period | | | Earnings in each period conditional on working in that period | | |
|---|---|---|---|---|---|---|
| | *T*+1 | *T*+2 | *T*+3 | *T*+1 | *T*+2 | *T*+3 |
| **Cancer** | | | | | | |
| Coefficient | -5,079.37 *** | -3,980.47 *** | -3,756.72 *** | -4,674.72 *** | -2,940.20 ** | -2,427.52 ** |
| Standard error | 902.42 | 878.38 | 810.23 | 953.02 | 939.40 | 880.57 |
| *n* (weighted sample) | 596,374 | 595,210 | 594,014 | 577,036 | 564,400 | 552,832 |
| **Survival category** | | | | | | |
| **High** | | | | | | |
| Coefficient | -4,420.21 ** | -3,264.05 * | -3,018.97 * | -4,134.35 ** | -1,953.90 | -1,788.80 |
| Standard error | 1,520.97 | 1,483.54 | 1,237.04 | 1,600.95 | 1,606.21 | 1,350.72 |
| **Middle** | | | | | | |
| Coefficient | -4,441.26 *** | -3,411.60 ** | -2,943.02 * | -3,951.73 *** | -2,971.60 ** | -2,335.15 † |
| Standard error | 1,135.93 | 1,066.17 | 1,230.35 | 1,185.57 | 1,099.84 | 1,290.12 |
| **Low** | | | | | | |
| Coefficient | -11,989.45 *** | -11,665.96 *** | -11,155.49 *** | -10,631.41 *** | -8,630.16 ** | -5,254.77 |
| Standard error | 2,475.29 | 2,509.91 | 2,744.28 | 2,774.28 | 2,818.35 | 3,493.13 |
| *n* (weighted sample) | 596,191 | 595,025 | 593,834 | 576,860 | 564,227 | 552,676 |

*** significantly different from reference category ($p<0.001$)

** significantly different from reference category ($p<0.01$)

* significantly different from reference category ($p<0.05$)

† significantly different from reference category ($p<0.10$)

**Note:** Results are estimates for a linear regression model with Coarsened Exact Matching (CEM) weights. All regressions conditional on the full set of control variables listed in Subsection 2.3. *n*: number of observations; *T*: year of cancer diagnosis for the cancer sample. Author's calculations.

**Source:** Statistics Canada, 1991 Census–Longitudinal Worker File.

## 4.3 Combining difference-in-differences with matching

Next, the study further controls for differences between the cancer and comparison samples by differencing out time-invariant unobserved characteristics that may be associated with both cancer incidence and earnings. Earnings at $T$-1 are subtracted from earnings at $T$+1, $T$+2, and $T$+3. These three pre- and post-diagnosis differences are then used as the dependent variables in the weighted DID regression. The top panel in Table 5 presents the results from the weighted DID regressions for annual earnings. For comparison, the middle panel in Table 5 shows again the results from the weighted linear regression reported in Table 4: the bottom panel shows the results from the linear regression based on the unweighted (pre-matching) sample.

Table 5 shows several notable patterns among the results from these three models.[25] First, in the top panel, after additionally accounting for unobservable differences between the cancer and comparison samples, the negative cancer effects become smaller compared with those in the middle panel, where only observable differences between the two samples are accounted for. These differences in the estimates provide evidence of modest time-invariant unobservable differences between the (matched) cancer and comparison samples: this leads to the overestimation of the negative cancer effect in the model reported in Table 4 (the middle panel of Table 5). Second, comparing the results in the middle panel to those in the bottom panel shows that, consistent with Ho et al. (2007), controlling for selection on observable characteristics by applying matching weights improves causal inference, particularly at $T$+1—a relatively short term in this study. In the longer term ($T$+3), however, accounting for selection on observable characteristics in the model appears less important. Third, accounting for the simultaneous associations between time-invariant unobservable individual characteristics and cancer, on the one hand, and these characteristics and labour market outcomes, on the other hand, appears more important for estimating longer-term effects of cancer on earnings than for estimating short-term effects.

25. Alternatively, Appendix Tables A.4-1 and A.4-2 present results using a sample with working at $T$ and $T$-1, and Appendix Tables A.5-1 and A.5-2 present those using a sample with working at $T$, $T$-1, $T$-2 and $T$-3. As can be expected, if the sample is restricted to those with stronger labour market attachment prior to the cancer diagnosis, the negative effect of cancer on earnings is likely to be larger but the differences are small. Overall results show the same patterns as in Table 5. There is no difference in the negative effect of cancer on employment regardless of the restrictions on individual's work status prior to the cancer diagnosis.

**Table 5**

**The effect of cancer on earnings (in dollars) controlling for unobservable individual characteristics**

| | Unconditional on working at *T+j* | | | Conditional on working at *T+j* | | |
|---|---|---|---|---|---|---|
| | *T*+1 | *T*+2 | *T*+3 | *T*+1 | *T*+2 | *T*+3 |
| **Weighted DID regression** | | | | | | |
| **Cancer** | | | | | | |
| Coefficient | -4,967.85 *** | -3,858.98 *** | -3,386.03 *** | -4,269.39 *** | -2,428.57 *** | -1,763.90 ** |
| Standard error | 780.07 | 748.05 | 735.03 | 703.10 | 625.02 | 639.44 |
| *n* (weighted sample) | 596,374 | 595,210 | 594,014 | 577,036 | 564,400 | 552,832 |
| **Weighted linear regression (upper panel of Table 4)** | | | | | | |
| **Cancer** | | | | | | |
| Coefficient | -5,079.37 *** | -3,980.47 *** | -3,756.72 *** | -4,674.72 *** | -2,940.20 ** | -2,427.52 ** |
| Standard error | 902.42 | 878.38 | 810.23 | 953.02 | 939.40 | 880.57 |
| *n* (weighted sample) | 596,374 | 595,210 | 594,014 | 577,036 | 564,400 | 552,832 |
| **Unweighted sample OLS** | | | | | | |
| **Cancer** | | | | | | |
| Coefficient | -5,233.87 *** | -4,076.90 *** | -3,821.77 *** | -4,833.49 *** | -3,047.78 ** | -2,472.27 ** |
| Standard error | 902.52 | 876.65 | 814.36 | 952.25 | 935.19 | 881.80 |
| *n* (unweighted sample) | 1,166,397 | 1,163,918 | 1,161,372 | 1,129,985 | 1,107,613 | 1,088,126 |

*** significantly different from reference category ($p<0.001$)

** significantly different from reference category ($p<0.01$)

* significantly different from reference category ($p<0.05$)

**Note:** All regressions conditional on the full set of control variables listed in Subsection 2.3. DID: difference-in-differences; *n*: number of observations; OLS: ordinary least squares; *T*: year of cancer diagnosis in the cancer sample. Author's calculations.

**Source:** Statistics Canada, 1991 Census–Longitudinal Worker File.

## 4.4 Heterogeneity of cancer effect

So far, the analysis has focused on the overall effect of cancer on specific labour market outcomes. The question now is whether cancer effects for subgroups with certain observable characteristics are different from the average cancer effect estimated for the whole sample.[26] To answer this, models with interactions between the cancer indicator and categorical variables for age, education and pre-diagnosis earnings are estimated. Interactions between each set of categories and the cancer indicator are included separately in each regression; all control variables from previous regressions are also included. First, group-specific average cancer effects are computed; then, differences between these effects and the overall average effect (shown in Tables 3 and 4) are tested for statistical significance. A model with the interaction terms is used instead of estimating separate models for each category in order to hold cancer incidence constant over all subgroups by using CEM weights obtained for the whole sample. Because the weighted cancer incidence is the same in the whole sample and in each subgroup (e.g., people aged 35 to 44, people with a university degree, etc.), these differences between estimated group-specific and full-sample effects can be interpreted as those reflecting differences in (pre-cancer) characteristics between the subgroups and the whole sample and not as differences in cancer incidence.

26. The study does not compare cancer effects for one sub-category with those for the baseline category, as the heterogeneity in pre-cancer characteristics between the cancer and comparison samples is deliberately removed by applying matching weights.

The differences in cancer effects ($\Delta CE$) between each subgroup ($[CE \mid x = 1]$) and the overall effect estimated on the whole sample ($[CE]$) are presented in Tables 6-1 and 6-2. The following are the test hypotheses:

$$[\ \hat{pr}(work \mid cancer = 1 \; and \; x = 1) - \hat{pr}(work \mid cancer = 0 \; and \; x = 1)] - [\hat{pr}(work \mid cancer = 1) - \hat{pr}(work \mid cancer = 0)] = [CE_{work} \mid x = 1] - [CE_{work}] = 0 \quad (1)$$

$$[\ E(earnings \mid cancer = 1 \; and \; x = 1) - E(earnings \mid cancer = 0 \; and \; x = 1)] - [E(earnings \mid cancer = 1) - E(earnings \mid cancer = 0)] = [CE_{earnings} \mid x = 1] - [CE_{earnings}] = 0 \quad (2)$$

Table 6-1 presents $\Delta CEs$ for the probability of working at period $T$+1 and $T$+3 as well as test results for Equation (1) for each set of sample characteristics. The cancer effects on the probability of working for younger age groups (25-to-34 and 35-to-44) are 1 to 2 percentage points higher at $T$+1 than the cancer effect estimated for the full sample (-0.030 in Table 3), where the average age is 48:[27] this difference is statistically significant at the 10% level. The $\Delta CEs$ for the two youngest age groups are larger and more statistically significant at $T$+3 than at $T$+1. However, the $\Delta CEs$ for older age groups (45-to-54 and 55-to-61) are not statistically significant in either period. This result suggests that, for those diagnosed with cancer at younger ages, the magnitude of the negative cancer effect on the probability of working is smaller than the estimated average effect for the whole sample, particularly in the long run.

27. In other words, at period $T$+1, the average probability of working for those aged 25-to-34 in the year of cancer diagnosis ($T$) is 1 percentage point (-0.030+ 0.021 = -0.009) lower than the average probability of working for their counterparts in the comparison sample.

**Table 6-1**
**Heterogeneity of cancer effects on work status, various sample characteristics**

| | Differences in cancer effects ($\Delta CE$) | |
|---|---|---|
| | percentage points | standard error |
| ***T*+1**[1] | | |
| **Age at *T*** | | |
| 25 to 34 | 0.021 † | 0.013 |
| 35 to 44 | 0.013 † | 0.007 |
| 45 to 54 | -0.005 | 0.006 |
| 55 to 61 | -0.012 | 0.010 |
| **Highest level of schooling** | | |
| No high school | -0.010 | 0.008 |
| High school, with/without trades certification | 0.008 | 0.006 |
| Postsecondary non-university | -0.015 | 0.011 |
| University degree | 0.020 † | 0.012 |
| **Quintiles of earnings at *T*-1** | | |
| Lowest | -0.042 *** | 0.012 |
| Second | -0.013 | 0.010 |
| Third | 0.023 ** | 0.009 |
| Fourth | 0.023 * | 0.009 |
| Highest | 0.023 ** | 0.009 |
| ***T*+3**[2] | | |
| **Age at *T*** | | |
| 25 to 34 | 0.036 * | 0.016 |
| 35 to 44 | 0.020 * | 0.010 |
| 45 to 54 | -0.012 | 0.009 |
| 55 to 61 | -0.012 | 0.013 |
| **Highest level of schooling** | | |
| No high school | -0.019 † | 0.011 |
| High school, with/without trades certification | 0.004 | 0.009 |
| Postsecondary non-university | 0.009 | 0.014 |
| University degree | 0.017 | 0.016 |
| **Quintiles of earnings at *T*-1** | | |
| Lowest | -0.013 | 0.014 |
| Second | -0.019 | 0.013 |
| Third | 0.010 | 0.013 |
| Fourth | 0.011 | 0.014 |
| Highest | 0.017 | 0.013 |

*** significantly different from reference category ($p<0.001$)

** significantly different from reference category ($p<0.01$)

* significantly different from reference category ($p<0.05$)

† significantly different from reference category ($p<0.10$)

1. Sample average effect at *T*+1, -0.030.

2. Sample average effect at *T*+3, -0.048.

**Note:** Results are computed based on the probit model with Coarsened Exact Matching (CEM) weights. To estimate the age-group-specific cancer effects, age categories are included instead of age and age squared. Otherwise, all regressions are conditional on interactions between cancer and a set of the categorical variables and the full set of control variables listed in Subsection 2.3.

*T*: year of cancer diagnosis for the cancer sample. *T*-1: *T* minus one. Author's calculations.

**Source:** Statistics Canada, 1991 Census–Longitudinal Worker File.

The $\Delta CEs$ are also statistically significant in some educational categories. At $T$+1 the only significant $\Delta CE$ is observed for cancer survivors with a university degree for whom the cancer effect on the probability of working is 2 percentage points higher than the overall effect. At $T$+3, the only significant $\Delta CE$ is observed for those with no high school diploma; it is 1.9 percentage points lower than the overall cancer effect on the probability of working at $T$+3. In other words, at $T$+3 the negative effect of cancer on the employment status among individuals with no high school diploma is larger in magnitude (-0.048 - 0.019 = -0.067) than the overall average effect (-0.048).[28]

For cancer survivors in the bottom 20% of the earning distribution at $T$-1, the $\Delta CE$ at $T$+1 is -0.042, which means that the negative effect is 4.2 percentage points larger than the overall effect (-0.030). In contrast, for those in the upper 60% of the earnings distribution at $T$-1, the $\Delta CE$ at $T$+1 (0.023) is positive and statistically significant, so the negative effects of cancer on the work status for these groups are close to zero at $T$+1 (-0.03+0.023= -0.007). At $T$+3, the corresponding $\Delta CEs$ are much smaller and neither is statistically significant.

Table 6-2 presents $\Delta CEs$ for annual earnings and test results from Equation (2) at $T$+1 and $T$+3. Post-cancer earnings losses are generally small for cancer survivors in the 25-to-34 age group (-5,079.37+4,337.53=-741.87) for which the negative effect on the probability of working is smaller relative to the overall average effect. However, the patterns of group-specific cancer effects on earnings are less clear than those observed for work status in Table 6-1.

28. To a certain degree, it may because the people with no high school diploma are more likely to be diagnosed with cancer with low survival rate (or more severe cancer) such as lung cancer. In general, there is a negative relationship between level of education and smoking behavior (Maralani 2013).

## Table 6-2
### Heterogeneity of cancer effects on earnings (zero earnings included), by sample characteristics

| | Differences in cancer effects ($\Delta CE$) | |
|---|---|---|
| | dollars | standard error |
| *T*+1[1] | | |
| **Age at *T*** | | |
| 25 to 34 | 4,337.53 ** | 1,581.50 |
| 35 to 44 | -982.43 | 1,098.64 |
| 45 to 54 | -1,361.36 | 1,035.92 |
| 55 to 61 | 1,700.80 | 1,991.31 |
| **Highest level of schooling** | | |
| No high school | 400.24 | 1,165.61 |
| High school, with/without trades certification | 149.84 | 995.47 |
| Postsecondary non-university | -903.13 | 1,417.18 |
| University degree | 78.07 | 3,881.43 |
| **Quintiles of earnings at *T*-1** | | |
| Lowest | 2,802.57 ** | 1,064.16 |
| Second | 666.51 | 1,024.41 |
| Third | -43.56 | 1,082.79 |
| Fourth | -762.02 | 1,206.78 |
| Highest | -3,674.27 | 3,833.93 |
| *T*+3[2] | | |
| **Age at *T*** | | |
| 25 to 34 | 5,020.94 ** | 1,740.39 |
| 35 to 44 | -1,181.10 | 1,033.17 |
| 45 to 54 | -720.37 | 964.59 |
| 55 to 61 | 752.23 | 1,639.10 |
| **Highest level of schooling** | | |
| No high school | -597.08 | 1,022.11 |
| High school, with/without trades certification | 733.24 | 920.51 |
| Postsecondary non-university | 234.04 | 1,334.48 |
| University degree | -1,130.42 | 3,287.29 |
| **Quintiles of earnings at *T*-1** | | |
| Lowest | 2,081.94 * | 977.97 |
| Second | 1,322.43 | 954.97 |
| Third | 849.97 | 1,079.76 |
| Fourth | -226.33 | 1,284.24 |
| Highest | -5,088.98 | 3,292.04 |

** significantly different from reference category ($p<0.01$)

* significantly different from reference category ($p<0.05$)

1. Sample average effect at *T*+1, -5,079.37.

2. Sample average effect at *T*+3, -3,756.72.

**Note:** Results are computed based on a linear regression model with Coarsened Exact Matching (CEM) weights. To estimate the age-group-specific cancer effects, age categories are included instead of age and age squared. Otherwise, all regressions are conditional on interactions between cancer and a set of the categorical variables and the full set of control variables listed in Subsection 2.3. *T*: year of cancer diagnosis for cancer sample. *T*-1: *T* minus one. Author's calculations.

**Source:** Statistics Canada, 1991 Census–Longitudinal Worker File.

## 5 Conclusion

Earlier diagnoses, improvements in cancer treatment and better follow-up care have substantially increased the number of cancer survivors during the last two decades. As cancer survival rates rise, life after cancer treatment has attracted considerable interest of researchers in different fields. Economists are particularly interested in the effects of cancer on the labour market choices and labour market outcomes of cancer survivors.

Using unique linkage data from the Canadian 1991 Census and administrative cancer, mortality and longitudinal income tax records, the study estimates the effects of cancer on certain labour market outcomes of cancer survivors by comparing their outcomes to those of a comparison group consisting of people never diagnosed with cancer. On average, cancer lowers the probability of working in the first year following the year of diagnosis by 3 percentage points relative to the comparison sample. The average earnings loss of cancer survivors in the first post-diagnosis year is 12% when transitions from employment to non-employment are taken into account by including zero earners in the sample. When only those with non-zero annual earnings are considered, the estimated average earnings loss for cancer survivors is 10% to 11%. In the second and third year after the year of diagnosis, cancer lowers the probability of working among cancer survivors; however, the earnings gap between survivors and those in the comparison group narrows, particularly when cancer survivors continue to work after the cancer diagnosis. In addition to controlling for differences in observable characteristics between people diagnosed with cancer and those never diagnosed with cancer, accounting for the simultaneous associations between (1) time-invariant unobservable individual characteristics and cancer, and (2) between these characteristics and labour market outcomes appears important for estimating long-term effects of cancer on earnings.

The study findings also suggest that cancer effects on labour market outcomes differ for different survival categories, which in this study serve as proxies for cancer severity. The probability of working and annual earnings are substantially lower for survivors in the low-survival category than for those in the high-survival category in all post-cancer periods. In the long run, the negative effect of cancer on the work status of people diagnosed at ages younger than the average age in the sample is smaller than the average effect for the full cancer sample. Conversely, the negative effect of cancer on the work status of cancer survivors with no high school diploma is greater than the average effect for the full sample.

Overall, the findings suggest that, in the long run, cancer is more likely to affect survivors' work status (extensive margin of labour market outcomes) than their earnings (intensive margin of labour market outcomes). Improving general understanding of the role that supply and demand sides of the labour market play in reducing the likelihood of employment of cancer survivors relative to those never diagnosed with cancer is an important direction for future research.

# 6 Appendix

## Table A.1
## Cancer sites, cancer survival category and death records

| | Cancer site | | Cancer survival category | Death records found by 2006 within each group |
|---|---|---|---|---|
| | Number of persons | Share | | |
| | number | | percent | |
| **High-survival category** | ... | ... | 39.02 | ... |
| Thyroid | 138 | 2.66 | ... | ... |
| Prostate | 305 | 5.88 | ... | ... |
| Testis | 56 | 1.08 | ... | ... |
| Skin melanoma | 241 | 4.65 | ... | ... |
| Breast | 1,117 | 21.54 | ... | ... |
| Corpus uteri | 136 | 2.62 | ... | ... |
| Hodgkin lymphoma | 30 | 0.58 | ... | ... |
| Deaths | ... | ... | ... | 19.13 |
| **Middle-survival category** | ... | ... | 32.00 | ... |
| Chronic lymphocytic leukemia | 32 | 0.62 | ... | ... |
| Cervix uteri | 409 | 7.89 | ... | ... |
| Bladder (including in situ) | 147 | 2.84 | ... | ... |
| Kidney and renal pelvis | 145 | 2.80 | ... | ... |
| Soft tissue | 39 | 0.75 | ... | ... |
| Larynx | 51 | 0.98 | ... | ... |
| Rectum | 177 | 3.41 | ... | ... |
| Colon | 298 | 5.75 | ... | ... |
| Non-Hodgkin lymphoma | 229 | 4.42 | ... | ... |
| Oral (buccal cavity and pharynx) | 132 | 2.55 | ... | ... |
| Deaths | ... | ... | ... | 34.30 |
| **Low-survival category** | ... | ... | 21.04 | ... |
| Ovary | 108 | 2.08 | ... | ... |
| Multiple myeloma | 43 | 0.83 | ... | ... |
| Leukemia (excluding chronic lymphocytic leukemia) | 71 | 1.37 | ... | ... |
| Stomach | 74 | 1.43 | ... | ... |
| Brain | 108 | 2.08 | ... | ... |
| Liver | 35 | 0.68 | ... | ... |
| Lung and bronchus | 520 | 10.03 | ... | ... |
| Esophagus | 37 | 0.71 | ... | ... |
| Pancreas | 95 | 1.83 | ... | ... |
| Deaths | ... | ... | ... | 81.48 |
| **Missing** | ... | ... | 7.95 | ... |
| Others | 412 | 7.95 | ... | ... |
| Deaths | ... | ... | ... | 42.48 |
| **Total** | 5,185 | 100 | 100 | ... |

... not applicable

**Note:** The 1991 Census–Longitudinal Worker File data contain a total of 5,185 individuals diagnosed with cancer for the first time between 1992 and 2000 and aged 61 and younger in the year of the diagnosis. Cancer sites are defined based on Surveillance, Epidemiology, and End Results (SEER) grouping definitions for the International Classification of Diseases for Oncology, Second and Third Edition (ICD-O-2/3). Figures may not add up to totals because of rounding. Author's calculations.

**Source:** Statistics Canada, 1991 Census–Longitudinal Worker File.

**Table A.2**
**Summary statistics for other characteristics of the cancer and comparison group before and after matching and pruning**

| Variables[1] | Cancer sample | Comparison sample: Unmatched | Comparison sample: Matched and reweighted | Unmatched difference[2] |
|---|---|---|---|---|
| | | dollars | | |
| Earnings at *T*-1 (mean) | 43,445 | 46,454 | 43,374 | -3,009 * |
| | | percent | | percentage points |
| **Quintiles of earnings at *T*-1** | | | | |
| Lowest | 23.45 | 19.99 | 23.10 | 3.46 *** |
| Second | 21.79 | 20.00 | 21.64 | 1.80 * |
| Third | 19.25 | 20.00 | 20.02 | -0.75 |
| Fourth | 17.60 | 20.01 | 17.85 | -2.41 ** |
| Highest | 17.91 | 20.00 | 17.39 | -2.10 ** |
| | | dollars | | |
| Earnings at *T*-2 (mean) | 43,689 | 45,493 | 42,902 | -1,804 |
| | | percent | | percentage points |
| **Quintiles of earnings at *T*-2** | | | | |
| Lowest | 24.37 | 19.99 | 22.84 | 4.38 *** |
| Second | 20.83 | 20.00 | 21.32 | 0.83 |
| Third | 18.75 | 20.00 | 20.18 | -1.25 |
| Fourth | 17.10 | 20.01 | 17.92 | -2.91 *** |
| Highest | 18.95 | 20.00 | 17.75 | -1.06 |
| **Self-employment at *T*** | | | | |
| No | 82.13 | 80.54 | 80.62 | 1.60 * |
| Yes | 15.94 | 16.73 | 17.14 | -0.79 |
| Missing | 1.93 | 2.74 | 2.25 | -0.81 * |
| Union at T | 41.16 | 40.87 | 40.82 | 0.29 |
| **Marital status at *T*** | | | | |
| Single | 23.57 | 22.92 | 22.37 | 0.64 |
| Couple | 73.55 | 72.96 | 74.18 | 0.59 |
| Missing | 2.89 | 4.12 | 3.45 | -1.23 ** |
| **Long-term disabilities or handicaps in 1991 Census** | | | | |
| No | 96.65 | 96.81 | 96.53 | -0.16 |
| Yes | 2.43 | 2.25 | 2.56 | 0.17 |
| Missing | 0.92 | 0.94 | 0.91 | -0.01 |
| | | number | | |
| *n* | 2,597 | 1,228,551 | 624,835 | ... |

... not applicable

*** significantly different from reference category (p<0.001)

** significantly different from reference category (p<0.01)

* significantly different from reference category (p<0.05)

1. Variables not used in matching but included in the regression analysis.

2. This column displays the difference in means (proportion) between cancer and comparison group before matching; standard symbols indicate significant t-test differences between these two groups.

**Note:** Earnings in *T*-2 (*T* minus 2) are not included in regression analysis. *T*: year of cancer diagnosis for the cancer sample; *n*: number of observations. *T*-1: *T* minus 1. Author's calculations.

**Source:** Statistics Canada, 1991 Census–Longitudinal Worker File.

### Table A.3
### Unadjusted mean of employment and earnings in period $T$+1, $T$+2 and $T$+3

| | $T$+1 | $T$+2 | $T$+3 |
|---|---|---|---|
| **Cancer sample** | | | |
| **All cancer survivors** | | | |
| Working (percent) | 92.38 | 89.03 | 85.28 |
| Earnings (mean dollars) | 37,124 | 37,517 | 36,577 |
| *n* | 2,573 | 2,561 | 2,540 |
| **High survival category** | | | |
| Working (percent) | 93.26 | 88.59 | 85.27 |
| Earnings (mean dollars) | 38,149 | 38,412 | 37,266 |
| *n* | 1,291 | 1,288 | 1,276 |
| **Middle survival category** | | | |
| Working (percent) | 92.10 | 90.81 | 88.66 |
| Earnings (mean dollars) | 37,424 | 37,891 | 37,377 |
| *n* | 886 | 881 | 873 |
| **Low survival category** | | | |
| Working (percent) | 84.47 | 79.70 | 69.80 |
| Earnings (mean dollars) | 29,350 | 29,179 | 28,517 |
| *n* | 206 | 202 | 202 |
| **Missing survival category** | | | |
| Working (percent) | 96.32 | 93.68 | 86.24 |
| Earnings (mean dollars) | 37,182 | 38,579 | 36,837 |
| *n* | 190 | 190 | 189 |
| **Comparison sample** | | | |
| **Unmatched non-cancer** | | | |
| Working (percent) | 96.85 | 95.19 | 93.77 |
| Earnings (mean dollars) | 47,145 | 47,296 | 47,402 |
| *n* | 1,220,406 | 1,214,199 | 1,209,296 |
| **Matched non-cancer** | | | |
| Working (percent) | 96.75 | 94.87 | 93.19 |
| Earnings (mean dollars) | 46,146 | 45,933 | 45,691 |
| *n* | 621,050 | 617,958 | 615,693 |

**Note:** Calculations of mean earnings include zero earnings of people who did not work during the period.
*n*: number of observations; *T*: year of cancer diagnosis for the cancer sample. Author's calculations.
**Source:** Statistics Canada, 1991 Census–Longitudinal Worker File.

### Table A.4-1
### The effect of cancer on employment for the sample restricted with working at *T* and *T*-1

| | Work in each period: Pr($work_{T+j}$=1) | | | Work continually up to each period: Pr($work_{T+j}$=1 \| $work_{T+j-1}$=1, $work_{T+j-2}$=1) | | |
|---|---|---|---|---|---|---|
| | *T*+1 | *T*+2 | *T*+3 | *T*+1 | *T*+2 | *T*+3 |
| **Cancer** | | | | | | |
| Coefficient | -0.033 *** | -0.037 *** | -0.049 *** | -0.033 *** | -0.017 *** | -0.014 ** |
| Standard error | 0.005 | 0.006 | 0.007 | 0.005 | 0.005 | 0.005 |
| *n* (weighted sample) | 612,760 | 610,639 | 608,809 | 612,760 | 589,413 | 567,726 |

*** significantly different from reference category (p<0.001)

** significantly different from reference category (p<0.01)

**Note:** This table replicates the upper panel of Table 3 for the sample restricted to individuals working at *T* and *T*-1 (*T* minus 1). The matched sample sizes at *T* are 2,671 for a cancer and 651,506 for a comparison sample. *n:* number of observations; *T*: year of cancer diagnosis for cancer sample. Author's calculations.

**Source:** Statistics Canada, 1991 Census–Longitudinal Worker File.

### Table A.4-2
### The effect of cancer on earnings (in dollars) for the sample restricted with working at *T* and *T*-1

| | Unconditional on working at *T+j* | | | Conditional on working at *T+j* | | |
|---|---|---|---|---|---|---|
| | *T*+1 | *T*+2 | *T*+3 | *T*+1 | *T*+2 | *T*+3 |
| **Weighted DID regression** | | | | | | |
| **Cancer** | | | | | | |
| Coefficient | -4,820.26 *** | -3,768.02 *** | -3,297.79 *** | -4,132.46 *** | -2,389.55 *** | -1,710.38 ** |
| Standard error | 759.68 | 728.78 | 717.45 | 689.12 | 612.51 | 628.43 |
| *n* (weighted sample) | 621,028 | 619,714 | 618,473 | 599,218 | 585,719 | 573,801 |
| **Weighted linear regression** | | | | | | |
| **Cancer** | | | | | | |
| Coefficient | -4,977.03 *** | -3,911.01 *** | -3,715.15 *** | -4,577.64 *** | -2,894.82 ** | -2,364.64 ** |
| Standard error | 881.84 | 859.15 | 794.58 | 936.67 | 921.15 | 865.34 |
| *n* (weighted sample) | 621,028 | 619,714 | 618,473 | 599,218 | 585,719 | 573,801 |
| **Unweighted sample OLS** | | | | | | |
| **Cancer** | | | | | | |
| Coefficient | -5,117.83 *** | -4,004.28 *** | -3,760.47 *** | -4,728.78 *** | -2,998.04 ** | -2,396.81 ** |
| Standard error | 882.29 | 858.03 | 798.40 | 936.18 | 917.65 | 866.40 |
| *n* (unweighted sample) | 1,202,688 | 1,199,828 | 1,197,110 | 1,161,723 | 1,137,848 | 1,117,645 |

*** significantly different from reference category (p<0.001)

** significantly different from reference category (p<0.01)

**Note:** This table replicates Table 5 for the sample restricted to individuals working at *T* and *T*-1 (*T* minus 1). The matched sample sizes at *T* are 2,671 for a cancer and 651,506 for a comparison sample. DID: difference-in-differences; *n:* number of observations; OLS: ordinary least squares; *T*: year of cancer diagnosis for the cancer sample. Author's calculations.

**Source:** Statistics Canada, 1991 Census–Longitudinal Worker File.

## Table A.5-1

## The effect of cancer on employment for the sample restricted with working at *T*, *T*-1, *T*-2 and *T*-3

| | Work in each period: Pr($work_{T+j}$=1) | | | Work continually up to each period: Pr($work_{T+j}$=1 \| $work_{T+j-1}$=1, $work_{T+j-2}$=1) | | |
|---|---|---|---|---|---|---|
| | *T+1* | *T+2* | *T+3* | *T+1* | *T+2* | *T+3* |
| **Cancer** | | | | | | |
| Coefficient | -0.031 *** | -0.038 *** | -0.049 *** | -0.031 *** | -0.019 *** | -0.014 ** |
| Standard error | 0.005 | 0.006 | 0.007 | 0.005 | 0.005 | 0.005 |
| *n* (weighted sample) | 567,291 | 565,602 | 564,007 | 567,291 | 548,372 | 529,543 |

*** significantly different from reference category (p<0.001)

** significantly different from reference category (p<0.01)

**Note:** This table replicates the upper panel of Table 3 for the sample restricted to individuals working at *T*, *T*-1 (*T* minus 1), *T*-2 (*T* minus 2) and *T*-3 (*T* minus 3). The matched sample sizes at *T* are 2,518 for a cancer and 601,590 for a comparison sample. *n*: number of observations; *T*: year of cancer diagnosis for cancer sample. Author's calculations.

**Source:** Statistics Canada, 1991 Census–Longitudinal Worker File.

## Table A.5-2

## The effect of cancer on earnings (in dollars) for the sample restricted with working at *T*, *T*-1, *T*-2 and *T*-3

| | Unconditional on working at *T+j* | | | Conditional on working at *T+j* | | |
|---|---|---|---|---|---|---|
| | *T+1* | *T+2* | *T+3* | *T+1* | *T+2* | *T+3* |
| **Weighted DID regression** | | | | | | |
| **Cancer** | | | | | | |
| Coefficient | -5,125.75 *** | -4,011.63 *** | -3,582.76 *** | -4,403.17 *** | -2,526.39 *** | -1,941.73 ** |
| Standard error | 801.48 | 767.53 | 749.09 | 721.18 | 640.12 | 647.78 |
| *n* (weighted sample) | 574,695 | 573,679 | 572,570 | 556,933 | 544,931 | 533,775 |
| **Weighted linear regression** | | | | | | |
| **Cancer** | | | | | | |
| Coefficient | -5,314.87 *** | -4,201.35 *** | -4,021.46 *** | -4,931.66 *** | -3,147.42 ** | -2,679.87 ** |
| Standard error | 923.75 | 899.01 | 824.45 | 975.52 | 962.15 | 896.13 |
| *n* (weighted sample) | 574,695 | 573,679 | 572,570 | 556,933 | 544,931 | 533,775 |
| **Unweighted sample OLS** | | | | | | |
| **Cancer** | | | | | | |
| Coefficient | -5,449.73 *** | -4,333.37 *** | -4,120.28 *** | -5,062.16 *** | -3,283.89 *** | -2,756.85 ** |
| Standard error | 922.70 | 896.64 | 828.60 | 973.46 | 957.01 | 896.97 |
| *n* (unweighted sample) | 1,134,030 | 1,131,814 | 1,129,527 | 1,100,455 | 1,079,254 | 1,060,522 |

*** significantly different from reference category (p<0.001)

** significantly different from reference category (p<0.01)

**Note:** This table replicates Table 5 for the sample restricted to individuals working at *T*, *T*-1 (*T* minus one), *T*-2 (*T* minus 2) and *T*-3 (*T* minus 3). The matched sample sizes at *T* are 2,518 for a cancer and 601,590 for a comparison sample. DID: difference-in-differences; *n*: number of observations; OLS: ordinary least squares; *T*: year of cancer diagnosis for cancer sample. Author's calculations.

**Source:** Statistics Canada, 1991 Census–Longitudinal Worker File.

## References

American Cancer Society. 2012. *Cancer Facts & Figures 2012.* Atlanta: American Cancer Society.

Bradley, C.J., H.L. Bednarek, and D. Neumark. 2002. "Breast cancer survival, work and earnings." *Journal of Health Economics* 21 (5): 757–779.

Bradley, C.J., D. Neumark, H.L. Bednarek, and M. Schenk. 2005. "Short-term effects of breast cancer on labour market attachment: results from a longitudinal study." *Journal of Health Economics* 24 (1): 137–160.

Bradley, C.J., D. Neumark, Z. Luo, and M. Schenk. 2007. "Employment and cancer: findings from a longitudinal study of breast and prostate cancer survivors." *Cancer Investigations* 25 (1): 47–54.

Carpenter, M., M.E. Fair, C. Poliquin, and P. Lalonde. 2008. *Canadian Cancer Data Base, 1969 to 1991 History and Development.* Occupational and Environmental Health Research Section, Report no. 16. Health Statistics Division. Ottawa: Statistics Canada.

Canadian Cancer Society (CCS). 2012. *Canadian Cancer Statistics 2012.* Canadian Cancer Society's Steering Committee on Cancer Statistics. Toronto: Canadian Cancer Society.

Centers for Disease Control and Prevention. 2011. "Cancer survivors—United States, 2007." *Morbidity and Mortality Weekly Report 2011* 60: 269–272.

Ellison, L.F., H. Bryant, G. Lockwood, and L. Shack. 2011. "Conditional survival analyses across cancer sites." *Health Reports* 22 (2): 21–25. Statistics Canada Catalogue no. 82-003-XPE.

Heckman, J., H. Ichimura, J. Smith, and P. Todd. 1998. "Characterizing selection bias using experimental data." *Econometrica* 66 (5): 1017–1098.

Heckman, J., H. Ichimura, and P. Todd. 1997. "Matching as an econometric evaluation estimator." *Review of Economic Studies* 65: 261–294.

Ho, D., K. Imai, G. King, and E. Stuart. 2007. "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference." *Political Analysis* 15 (3): 199–236.

Iacus, S.M., G. King, and G. Porro. 2011. "Multivariate matching methods that are monotonic imbalance bounding." *Journal of the American Statistical Association* 106 (493): 345–361.

Iacus, S.M., G. King, and G. Porro. 2012. "Causal inference without balance checking: Coarsened Exact Matching." *Political Analysis* 20 (1): 1–24.

Imbens, G.W., and J.M. Wooldridge. 2009. "Recent developments in the econometrics of program evaluation." *Journal of Economic Literature* 47 (1): 5–86.

Jones, A.M., N. Rice, and P. Rosa Dias. 2011. "Long-term effects of school quality on health and lifestyle: evidence from comprehensive schooling reforms in England." *Journal of Human Capital* 5 (3): 342–376.

Maralani, V. 2013. "Educational inequalities in smoking: The role of initiation versus quitting." *Social Science & Medicine* 84: 129–137.

Moran, J.R., P.F. Short, and C.S. Hollenbeak. 2011. "Long-term employment effects of surviving cancer." *Journal of Health Economics* 30 (3): 505–514.

Organisation for Economic Co-operation and Development (OECD). 2007. *Health at a Glance 2007: OECD indicators*. Paris: OECD Publishing.

Statistics Canada. 2008. *Canadian Cancer Registry System Guide, 2007 Edition*. Canadian Cancer Registry Redesign Team. Statistics Canada Catalogue no. 82-225-X. Ottawa: Statistics Canada.

Wilkins, R., M. Tjepkema, C. Mustard, and R. Choinière. 2008. "The Canadian census mortality follow-up study, 1991 through 2001." *Health Reports* 19 (3): 25–43. Statistics Canada Catalogue no. 82-003-X. Ottawa: Statistics Canada.